Deriving Common Data Elements from Real-Word Data for Alzheimer’s Disease and Alzheimer’s Disease Related Dementias

Technical Expert Panel Meeting

Sponsor: National Institute on Aging
Contract No.: 75FCMC18D0047
Outcome No: 7

McLean, VA

Robert D. Lieberthal
Niyati H. Vakil
Melissa B. Wilson

February 2023

This page intentionally left blank.

Acknowledgments

We wish to acknowledge and thank the fourteen members of the expert panel that presented in this meeting. Their names are contained in the Appendix (Appendix A). Their feedback and participation constitute most of the content contained in this report.

We also gratefully acknowledge additional meeting support provided by Nils Granholm, Laura Homer, and Allison Ounanian from MITRE and Cooper Roache, Dana Carluccio, and Christina Huffman from Rose Li and Associates.

Table of Contents

List of Tables

This page intentionally left blank.

1 Executive Summary

1.1 Meeting Purpose

On February 6, 2023, the National Institute on Aging (NIA) convened an exploratory discussion with fourteen expert panelists in the fields of Alzheimer's disease (AD)/Alzheimer's disease-related dementias (ADRD) biomedical and clinical research, epidemiology, and data science. The goal of the meeting was to identify opportunities to accelerate Alzheimer's Disease and Alzheimer's Disease Related Dementias (AD/ADRD) research using real-world-data (RWD) sources. The meeting focused primarily on Common Data Element (CDE) methods that can be applied for harmonizing data contained in RWD including healthcare claims and electronic health records (EHRs).

The meeting purpose was to examine the application of CDEs using harmonized and reliable RWD sources to aid research in the field. There is a recognized need for research which is integral to the development of treatments and interventions for AD/ADRD. NIA convened the meeting to determine the best use of data sources from claims and EHR and their role in common data models (CDMs). The plan for the meeting primarily included four primary outcomes:

- Eliciting preliminary expert feedback on developing common data element (CDE) domains for data harmonization.
- Identifying AD/ADRD research questions that real-world data (RWD) and CDE development can address.
- Prioritizing next steps for CDE development and additional CDEs for upcoming AD/ADRD therapeutics and biomarkers.
- Identifying actionable research priorities for NIA's consideration and research questions that could advance the field of aging research.

1.2 High Level Meeting Summary

Throughout the first half of the meeting, panelists discussed benefits and disadvantages to existing approaches to data harmonization and CDMs, agreeing that mapping data to a standard such as Observational Medical Outcomes Partnership (OMOP) can help CDE development based on EHR data. In addition, panelists expressed challenges in using coding systems (e.g., International Classification of Diseases [ICD] and Current Procedural Terminology [CPT] codes) to develop CDEs from payer claims data, including issues with tracking the historical usage of codes and heterogeneity in AD/ADRD diagnosis. During discussion of other RWD sources outside of EHR and claims data, panelists emphasized the importance collecting data to detect mild cognitive impairment (MCI) and early dementia symptoms, while also maintaining patient privacy and considering the health disparities of underrepresented populations.

The second half of the meeting focused developing CDEs for AD/ADRD research in the context of the following three real-world use cases:

- Combining private and public payer data.
- Developing CDEs from RWD for non-pharmaceutical intervention studies.

- Implementing and refining synthetic control methods in AD/ADRD research using CDE-based RWD.

For use case 1, panelists discussed the difficulties in combining Medicare and Medicaid claims data due to interstate variation in Medicaid data collection format and coverage. In addition, private claims data often lack important data elements, creating further challenges in combining public and private payer data. Next, for use case 2, panelists expressed that developing CDEs from data on non-pharmaceutical interventions is difficult because this data often lacks consistency in detail. However, certain high-performing integrated delivery systems do provide non-pharmaceutical interventions to patients. Lastly, for use case 3, panelists agreed that using synthetic controls in AD/ADRD clinical trials poses risks compared to using real-word control groups and may not be approved by regulatory agencies. As synthetic controls enable researchers to conduct emulation and RWD trials when true placebo groups may not be possible, which can aid in hypothesis development, it would be valuable to explore these methods in the future.

1.3 Domains for AD/ADRD Data CDEs

"A common data element (CDE) refers to a data element that is common to multiple data sets across different studies, surveys, or registries."[1] Examples of existing harmonization efforts and interoperability initiatives highlight the progress that has been made in CDE development and use of CDMs in health and healthcare data in general.

The meeting primarily focused on deriving CDEs from payer claims data and electronic health records (EHRs), although panelists discussed other RWD data sources that may help researchers track dementia progression, including data from imaging, nursing home settings, caregiver data (e.g., caregiver demographic data and health outcomes), car insurance data (e.g., accident frequency), and financial/credit data.

The application of CDEs to AD/ADRD Research was assessed by the panel as part of the meeting. The full panel addressed the lack of homogeneity in data definitions including what constitutes a CDE definition. Another important problem is existing data elements that have not been mapped to a framework. There was consensus about leveraging the interoperability among these standards, using mapping systems that allow for integration. Using administrative data to identify people with cognitive problems to start addressing variation in existing diagnostic codes and combining "top down" (method selection and implementation) with "bottom up" (development of specific CDEs or variables) approaches is likely to generate impact in AD/ADRD research.

The panel also recommended the development of seven domains to organize RWD for AD/ADRD research: 1) patient and caregiver information, 2) disease characterization, 3) health assessment, 4) biomarkers and genomics, 5) treatment, 6) patient and caregiver outcomes, and 7) non-health data. In addition, having a single provider collect and make derived variables in these domains may ameliorate concerns about data quality and consistency if that provider is properly regulated, and leveraging EHR and/or claims from healthcare systems that serve low-income, minoritized populations to gain traction on CDE approaches to ensure inclusion of underserved populations would be a way to promote health equity using these domains.

1.4 Additional CDEs for Development and Inclusion in RWD

The panel recommended several additional CDEs that should be developed and included in RWD. These included CDEs for quality of life (QoL) variables, end of life planning data, and educational characteristics of patients and their caregivers. The panel also emphasized the need for CDEs to describe physician characteristics for those providing AD/ADRD treatment. In addition, the panel emphasized the need to prioritize CDE development given the amount of work involved.

The panel also addressed the issue of CDE update frequency, especially given the rate of discovery and innovation in AD/ADRD research. The development of new blood and cerebrospinal fluid (CSF) tests, imaging modalities including PET scans, and new biomarkers all highlighted the need to regularly update any CDEs that are developed. The panel recommended annual updates as an ideal and biannual update as a minimum to ensure that any CDE system maintains accuracy and currency for the field.

Additional types of data that could be developed as CDEs also were discussed during the meeting. These include functional assessment data including Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) data, and data that could separate dementia from mild cognitive impairment (MCI) diagnosis and encourage severity assessment. Underrepresented populations that are not insured also may lack claims data, so the use of economic data for care other than claims, such as payments to federally qualified healthcare centers and safety net hospitals, was discussed.

Finally, data linkage for RWD, especially those linkages that can be enhanced through CDEs, was highlighted. Those data include Medicare and Medicaid (dual eligible) populations, combining fee-for-service (FFS) and Medicare Advantage data, and harmonization at a broad level with code books and support for individual researchers all would remove burden from individual research groups to allow scientists to focus on discovery and innovation in AD/ADRD.

1.5 New Research Questions that can be Answered Using Harmonized RWD

The expert group also identified several impactful research questions that could be addressed using the harmonized RWD discussed during the panel. These included early diagnosis and treatment of MCI, development of methods for improved diagnosis of dementia using RWD, and risk factor studies, including genetic risk factors, for AD/ADRD. The use of artificial intelligence (AI) as a complement to CDE development also was highlighted given the range of new AI tools and techniques that are available for research. The use of the Health and Retirement Study (HRS) data to determine validity of AI algorithms, while beyond the scope of the RWD discussed during the meeting, is one such example. Finally, research questions that examine disease trajectory and that identify and explain regional race and ethnic differences with regards to subtypes of dementia and diagnostic accuracy and lack thereof were highlighted as two additional areas for research that would be possible with harmonized RWD.

2 Meeting Summary

2.1 Session 1: Introduction to Real-World Data for Alzheimer's Disease and Alzheimer's Disease Related Dementias Research

2.1.1 Panelist Selection

Fourteen panel members were identified and invited in advance of the meeting to represent a range of perspectives as clinicians, social and behavioral researchers, and data managers. Most panel members were experts in, and regular users of, real-world data (RWD), with a focus on healthcare claims and electronic health record (EHR) data for AD/ADRD research. They represented academia, research organizations, and large health system perspectives.

Additional areas of expertise included common data elements (CDE) development and usage, clinical and research expertise in AD/ADRD comorbidities, health equity research and AD/ADRD science that focused on under-represented and historically disadvantaged groups. Current or prior research funding from NIA and other federal agencies was used to assess panelists' AD/ADRD research activities but was not required. These selection criteria ensured a professionally diverse group as shown in section A.2 of Appendix A.

2.1.2 Meeting Purpose

Partha Bhattacharyya, PhD, Chief Data Officer, Office of Data Resources and Analytics (ODRA) and Program Director, Division of Behavioral and Social Research, National Institute on Aging; Robert Lieberthal, PhD, MITRE

This session began with a welcome to all attendees and thanks to all including several National Institute on Aging (NIA) staff members who helped organize this meeting.

The presented meeting purpose was to examine the application of CDEs using harmonized and reliable RWD sources to aid research in the field. The introduction highlighted the need for research which is integral to the development of treatments and interventions for AD/ADRD. NIA wants to determine the best use of data sources (e.g., from health insurance claims, and electronic health record or EHR) and their role in common data models or CDMs.

The priorities for the meeting for NIA primarily were:

- Eliciting preliminary expert feedback on developing CDE domains for data harmonization
- Assessing AD/ADRD research questions that RWD and CDEs can address
- Determining a priority list of next steps for CDE development and additional CDEs for upcoming AD/ADRD therapeutics and biomarkers
- Identifying actionable research priorities for NIA's consideration and research questions that could advance the field of aging research

The pipeline of drugs and biomarkers also highlights the importance of these approaches and likely need for CDE development soon.

2.1.3 Pre-meeting Poll Results

The introductory session continued with a discussion of pre-meeting poll results from the expert panel. Poll results primarily related to the major challenges associated with deriving CDEs from RWD, applicable approaches, new research questions related to CDEs, and additional topic areas requested for further discussion during the meeting.

Major challenges highlighted and prioritized include:

- How best to identify persons living with AD/ADRD, under detection of cognitive issues evaluated in the clinical setting
- Validating data across multiple RWD datasets and developing a "gold standard" for identifying likely AD/ADRD from existing RWD CDEs (e.g., ICD codes, medications, cognitive assessment scores)
- Addressing barriers to harmonizing datasets, including differing formats, research stakeholders, study purposes, and data collection time periods such as:
  - Measures across different time periods collected and maintained in different formats, by different stakeholders, for different purposes under different incentive structures
  - Harmonizing datasets is beyond the scope of any specific research group

Applicable approaches to CDE development and common data models included Neurological Disorders and Stroke (NINDS) CDEs and Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) approaches.

New research questions included alignment of measures with clinical trials, development of algorithms to flag high risk patients to undergo additional screening with ancillary providers, assessment of how AD/ADRD impacts the health and financial well-being of families, and measures of level disability or dependencies the person may living with. Feedback about preliminary domains for AD/ADRD research provided before the meeting included several domains that could be consolidated or modified, with the greatest number of votes for "health assessment". Additional areas assessed included:

1. Subpopulations and increasing reliance on non-health system data for AD/ADRD diagnosis and treatment
2. Health equity, particularly harmonizing race, ethnicity, language, disability, sexual orientation, and gender identity (e.g., RELD SOGI) data and the approaches required to collect that data
3. Missing data elements including underdiagnosed disease, caregiver burden, and residential setting
4. Proxy data responses and the constructs used to elicit proxy responses
5. Data administration such as data linkage, privacy concerns, inconsistent data administration, and non-standardized or inconsistent data collection instruments
6. Using RWD sources beyond claims and EHR
7. Addressing negative societal attitudes towards dementia

8. Addressing comorbidity of multiple conditions that present barriers to diagnosis and treatment.

9. Identifying clinically approved and consistent biomarkers for AD/ADRD

2.2 Session 2: Common Data Elements Development for Alzheimer's Disease and Alzheimer's Disease Related Dementias Research

Allen Leavens, MD, MPH, MS, MITRE; Alex Whittaker, MS, MITRE

2.2.1 CDE Models

2.2.1.1 CDE Development for AD/ADRD Research

Existing data harmonization approaches to identify an approach to CDE development based on current best practices and methodologies were presented. Topics for meeting discussion were designed to focus primarily on harmonization approaches and CDE development based on RWD collected from EHRs and payer claims data. Data collected from EHRs, and payer claims create multiple overlapping CDEs, including International Classification of Diseases (ICD) codes, Hierarchical Condition Categories (HCC) codes, Medicare Severity-Diagnosis Related Groups (MS-DRG) codes, Current Procedural Terminology (CPT) codes, and National Drug Codes (NDCs). Although these codes provide details about diagnosed diseases, administered procedures, and prescribed medications, the use of these codes can vary across EHRs and claims data.

Multiple research initiatives have already developed approaches to data harmonization and CDE development. Although these approaches vary across different initiatives, each approach can offer insight into the development of CDEs for AD/ADRD research. For panelist discussions, MITRE prioritized CDE approaches developed by NINDS and the OMOP CDM developed by the Observational Health Data Sciences and Informatics (OHDSI) program. NINDS offers a searchable catalog of CDEs for particular diseases and disorders to aid efforts in harmonizing neuroscience data but does not focus on integration within a broader data model. In contrast, the OMOP CDM is an open community standard for harmonizing observational data sources and can be implemented with other data standards. Other existing data standards include United States Core Data for Interoperability (USCDI) and Fast Health Interoperability Resource (FHIR), both of which are widely implemented among various institutions and broader national models for health data exchange.

2.2.1.2 Expert Panel Subgroup Response

Panelists discussed existing data harmonization models and standards that are most applicable to AD/ADRD research, as well as the benefits and disadvantages to developing CDEs based on payment coding systems. Adoption of interoperable and broad standards (e.g., FHIR) for harmonization can aid AD/ADRD CDE development by establishing consistency in data reporting. Because many metadata resources already map to OMOP standards and many EHR systems already map to FHIR, developing CDEs under broad standards can enhance data-sharing across a greater number of institutions. In addition, multiple AD cohort studies have already attempted to map data to OMOP standards to further standardize cohort data.

2.2.1.3 Full Expert Panel Q&A

Expert Panel Reaction – *Full Expert Panel*

Although broad data standards can enhance data sharing, developing CDEs based on coding systems can provide greater detail in data elements, such as gender identity (including nonbinary). Also, EHRs often contain data with more heterogeneity than data in payer coding systems. However, multiple panelists noted that coding systems can also produce heterogenous data depending on how the systems are implemented. For example, many individuals with abnormal cognition are miscoded under the ADRD ICD code due to the ADRD code being used for different diagnoses in primary clinics compared to gerontology clinics. In addition, usage of ICD codes changes over time with new diagnostic knowledge, further creating heterogeneity within coding systems data.

Both broad data standards and detailed coding systems bring advantages to data harmonization and developing CDEs for AD/ADRD research. Using a combination of broad standards and coding systems may help counteract the heterogeneity of many existing data harmonization and CDE development approaches. For example, CDE development can begin with identifying critical ICD codes for AD/ADRD and adhering the collection of data to one of the existing broad standards. This method allows researchers to rapidly begin implementing CDEs for research measures while leveraging interoperability of data between many institutions.

2.2.1.4 CDE Model Future Directions

- Use administrative data to identify people with cognitive problems to start addressing variation in existing diagnostic codes
- Combine "top down" (method selection and implementation) with "bottom up" (development of specific CDEs or variables) approaches
- Develop and utilize different CDE methods for different data such as clinical trials, health economic studies, and observational research
- Address and adjudicate dementia versus non-dementia diagnosis in high utilization settings such as the emergency department

2.2.2 CDE Domains

2.2.2.1 Preliminary CDE Domains for AD/ADRD Research

Prior to the meeting, MITRE conducted research through the NIH RePORTER and the NIH National Library of Medicine (e.g., ClinicalTrials.gov and PubMed.gov) to identify potentially relevant CDEs and CDE domains for AD/ADRD research. These domains need to be designed to capture real-world data (RWD) meaning "…data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."[2] MITRE identified general data elements not currently used in NIA clinical trials and explored current and required approaches to collecting AD/ADRD data. The six preliminary CDE domains identified and presented to the panel for discussion were: patient information, disease characterization, health assessment, genomics, treatment, and outcomes.

2.2.2.2 Expert Panel Subgroup Response

Panelists discussed the challenges in obtaining sufficient detail from EHRs and claims data, other sources of data that may be relevant to AD/ADRD research and approaches to avoid biases in research resulting from inherent biases in RWD.

2.2.2.3 Full Expert Panel Q&A

Obtaining Information from EHR and Claims Data

Many patient test results are manually scanned into EHRs rather than electronically uploaded, which presents difficulties for extracting relevant test results and other relevant data elements from EHRs. In addition, payer claims data often provides insufficient information about patients that have multiple causes of dementia. However, biomarkers can help specify patient diagnostic information and can be captured from other EHR and payer claims data outside of test results, such as medications that require biomarkers to prescribe. Panelists also agreed that non-genomic biomarkers (e.g., blood amyloid and tau protein levels) and direct-to-consumer genetic tests should be considered as data sources for CDE development. In addition, natural language processing (NLP) can help analyze written text from EHRs and obtain more specific data than analyzing ICD codes alone.

Other Relevant RWD Sources

Mild cognitive impairment (MCI) is a more prevalent diagnosis at earlier ages than AD/ADRD and can be important to early treatment intervention. Early indications of cognitive impairment and dementia can potentially be derived from other sources of data outside of the six proposed CDE domains, such as changes in credit score and frequency of driving accidents. These data can also provide a more robust phenotype of individuals with MCI or dementia. However, the process of developing CDEs should respect the privacy of individuals and highlight specific data sources that should *not* be considered for collection.

Avoiding Biases in Research

Algorithms (e.g., machine learning [ML]) that integrate and utilize large amounts of data can save time by identifying patterns and associations that are otherwise difficult to observe. However, these algorithms can have innate biases that can exacerbate health disparities among underrepresented populations. For example, many non-native English speakers may have lower performance scores than native English speakers on cognitive assessments administered in English. If primary language is not incorporated as a data element in studies of cognitive assessment scores, then this lower cognitive performance could perpetuate biases against immigrants who are non-native English speakers. Panelists suggested that AD/ADRD CDEs incorporate data elements that can identify potential biases, such as primary language, patient racial/ethnic background, financial income, and living arrangements (e.g., ZIP code and whether a caregiver is present).

2.2.2.4 CDE Domain Future Directions

- Having a single provider collect and make derived variables may ameliorate some of these concerns if that provider is properly regulated
- Leveraging EHR and/or claims from healthcare systems that serve low-income, minoritized populations to gain traction on CDE approaches to ensure inclusion of underserved populations

- Enhance the "genomics" domain to include biomarkers
- Determine methods to share address or nine-digit zip code with researchers in a privacy-preserving manner, including in data feeds from sources such as the Centers for Medicare and Medicaid Systems (CMS)
- Ensure a domain for "non-health" data is included

2.3 Session 3: Use Cases

2.3.1 AD/ADRD RWD CDE Use Case Description

Prior to the meeting, NIA selected three use cases for potential CDEs to align with AD/ADRD research and integrate multiple RWD sources and longitudinal outcomes for aging populations. These use cases can help enhance CDE development and assessment of long-term impact of CDEs for AD/ADRD research. The subgroup panel discussions focused on the feasibility of these use cases for the scientific community.

The presentations considered three specific use cases for pilot implementation of CDE approaches that would provide a scope of work that will allow for implementation of CDE methods in AD/ADRD research:

- Use Case 1: Combining private insurer and public payer data
- Use Case 2: Developing CDEs from RWD for non-pharmaceutical intervention studies
- Use Case 3: Synthetic control methods for AD/ADRD research using CDE-based RWD

Each of these use cases satisfies criteria including direct feasibility for AD/ADRD research, the availability of multiple data sets within the use case, and likely acceptability within the research community. Each use case also has limitations that were explored in detail in three parts of session 3.

2.3.2 Use Case 1 – Combining Private Insurer and Public Payer Data

Robert Lieberthal, PhD, MITRE

2.3.2.1 Use Case Description

Use case 1 outlined the development of CDEs by combining private and public payer datasets, such as health insurance claims, Medicare claims, fee-for-service Medicare claims, and enrollment and claims for Medicare and Medicaid Advantage. This use case focused on combining payer data to develop CDEs for three primary reasons: (1) private and public payer data currently lacks shared identifiers to combine data, (2) combining private and public payer data can improve insurance coverage for new AD treatments, (3) insurance claims are one of the primary RWD data sources for research. Limitations to this approach include a lack of clinical data from payer claims data, privacy concerns with Personally Identifiable information (PII) and Protected Health Information (PHI), and difficulty harmonizing public and private data (e.g., consolidating various formats and codes). Examples of potential CDEs for use case 1 are shown in Table B-1 of Appendix B.

2.3.2.2 Expert Panel Subgroup Response

Panelists discussed barriers in collecting data from public and private payer claims data, the importance of context when analyzing payer claims data, and the need to update ICD codes used in payer claims data.

2.3.2.3 Full Expert Panel Q&A

NIA staff asked about priorities from the full panel. The example of death data including the national death index (NDI) was highlighted, as data may lag but eventually achieves high degrees of accuracy. The panel discussed the applicability for RWD for specific types of studies based on research need (e.g., longitudinal study versus a clinical trial (now clinical trial can vary from behavioral trial to drug, etc.). The "one size fits all" approach may not fit different needs of studies. The panel discussed the need for data contractors to consider providing this data in a more consistent manner. They also discussed the potential for misuse of data once CDEs are provided, since harmonized data alone may not be sufficient without the expertise needed to use the data.

<u>Barriers to Collecting Public and Private Payer Claims Data</u>

When collecting data from public payer claims, researchers often have more difficulty collecting Medicaid claims data compared to Medicare claims data because data collection formats and coverage of services vary from state to state, unlike Medicare claims data. In addition, many Medicare parts, such as Part D data, lack detail about the duration of medication use and whether medication use stopped because of deprescription or lack of patient adherence. In contrast, private claims data often lacks important data elements, such as racial/ethnic background, and researchers often cannot access multiple data elements simultaneously because of privacy concerns (e.g., reidentification). Because of these challenges in collecting payer claims data, these data sources may not provide enough information to completely populate AD/ADRD CDEs.

<u>Updating Payer Claims Data</u>

Panelists agreed that the use, limitations, and inaccuracies of CDEs in AD/ADRD research should be clear to researchers to prevent misinterpretation of data. Many researchers incorrectly implement algorithms to analyze large data sources without understanding the potential inaccuracies in data, such as limitations to ICD diagnostic codes (e.g., misdiagnosis of AD and ADRD). Panelists emphasized the importance of ensuring that AD/ADRD CDE analysis approaches that combine RWD sources do not replicate the potential for misinterpretation. To prevent data misinterpretation, payer claims data captured in CDEs should be updated to reflect changes in payment codes and ensure consistency of CDE data over time. In addition, clarifying the purpose of each CDE can help determine the quality and quantity of data needed to provide researchers with confidence regarding accurate analyses.

2.3.2.4 Use Case 1 Future Directions

- Regular updates of codes used in claims data
- Implementation of linkage strategies across health insurance providers
- Enhance and validate the accuracy of information to assess eligibility and outcome for drugs in prescription claims data

- Contractors or data providers should perform additional data cleaning to reduce the time and effort needed for individual research groups to work with claims data

2.3.3 Use Case 2 – Developing CDEs from RWD for Non-Pharmaceutical Intervention Studies

Allen Leavens, MD, MPH, MS, MITRE

2.3.3.1 Use Case Description

Use Case 2 outlined the development of CDEs for non-pharmaceutical interventions studies using both EHR and payer claims data, which can provide detailed data on large cohorts and observational studies and thus increase the impact and breadth of AD/ADRD research. For example, large-scale analyses of CDEs from RWD can help evaluate best practices for dementia care (e.g., pain management, mental health, and long-term care) and minimize the disruption and burden on patients by reducing the need for in-person studies. Limitations to this approach in developing CDEs include the inability to randomize individuals to specific non-pharmaceutical interventions and data privacy concerns. For this use case, all six of the proposed CDE domains are potentially applicable. However, along with payer claims data, EHRs do not consistently include data for relevant data elements because clinicians often have limited time with patients to capture data elements. Examples of potential CDEs for use case 2 are shown in Table B-2 of Appendix B.

2.3.3.2 Expert Panel Subgroup Response

The panelists discussed issues with capturing non-pharmaceutical intervention data and expanding the genomics CDE domain to include other risk factor data elements.

2.3.3.3 Full Expert Panel Q&A

Potential research opportunities include detecting rare outcomes, pathologies, and defining and examining various types of exposures (e.g., time-varying) and interventions (e.g., pharmaceutical) in addition to comparative effectiveness research. The experts also discussed the importance of CDEs tailored to the type of study and RWD (e.g., longitudinal, non-pharmaceutical interventions). For example, a historical data dictionary may be more relevant for longitudinal studies and less relevant to interventional studies. Useful domains include lifestyle and non-genetic risk factors. AI and natural language processing (NLP) for harmonization of data and the use of broader and narrower sets of CDEs (two-tiered, "top down" and "bottom up" approach) was raised.

Barriers to Capturing Non-Pharmaceutical Intervention Data

Many non-pharmaceutical interventions are conducted in non-clinical settings or in-home by caregivers, which poses challenges in collecting data from RWD sources. In addition, data collected on these interventions in clinical settings often provide inconsistent detail. For example, acupuncture therapy may provide data on exact pressure points targeted during a session, or no further data than the duration of the session. However, certain high performing health systems may require medical centers to implement non-pharmaceutical interventions and collect data, which can help researchers begin linking this data to AD/ADRD data elements. The VA has developed the Centralized Interactive Phenomics Resource (CIPHER), which provides definitions for both research clinical phenotypes and could benefit development of CDEs from RWD.

Panelists also noted the challenges in using CDEs for longitudinal observational and non-pharmaceutical intervention studies, reiterating that past data would need to be updated to reflect the changes in usage of codes.

<u>Expanding the Genomics CDE Domain</u>

Panelists also discussed challenges associated with limiting the genomics CDE domain to only genomic data elements. Genes associated with an individual developing AD, such as APOE ε4, may have high odds ratios for predicting AD/ADRD, but are often not recommended for consideration in clinical use because they are not sufficiently predictive without other diagnostic tools. The panelists suggested expanding the genomics CDE domain to include a broader scope of data elements, which can be analyzed for genetic and non-genetic risk factors. An additional risk factor domain can include lifestyle data elements, such as smoking, and other -omics test results, such as blood and CSF screenings for amyloid and tau proteins. However, data elements that capture blood and CSF screenings may be lower priorities for CDE development compared to more established CDEs because these screenings have not been validated for diagnostic use in primary care settings.

2.3.3.4 Use Case 2 Future Directions

- Broaden the genomics domain to include risk factors and biomarkers as a "biomarkers and genomics" domain
- Create a summary file for CMS FFS that is available for MA researchers or create one summary file of MA and FFS with a flag to indicate FFS or MA enrollment

2.3.4 Use Case 3 – Synthetic Control Methods for AD/ADRD Research using CDE-based RWD

Alex Whittaker, MS, MITRE

2.3.4.1 Use Case Description

Use Case 3 outlined developing CDEs for the creation of synthetic controls, which have been implemented in clinical trials and aggregate level interventions. Synthetic controls can enable researchers to conduct studies when true placebo controls are not feasible for ethical or logistical reasons. Synthetic controls also can increase the sample size of existing datasets, reduce recruitment needs for true controls, and enable alternative randomization approaches for studies where traditional randomization is challenging. Limitations to this approach in developing CDEs include the exclusion of underrepresented groups if data is not available, a lack of randomization for observational data, and a lack of regulatory acceptance. Also, developing CDEs for synthetic controls depends on study specifics (e.g., design or intervention type), which can complicate identifying CDEs. Although all six proposed CDE domains are potentially applicable, researchers creating synthetic controls will likely focus on CDEs linked to therapeutic interventions, genes associated with early onset AD, cognitive outcomes, and patient-reported outcomes. Examples of potential CDEs for use case 3 are shown in Table B-3 of Appendix B.

Medicare Part D claims data present the opportunity for researchers creating synthetic controls for therapeutic interventions to evaluate disease severity, adverse events, and diagnostics. Prescriptions for aducanumab and lecanemab are often captured in Medicare Part D claims, providing an opportunity for CDE development. NIA is also currently funding eight late-stage clinical trials for AD/ADRD drugs, which include therapeutics targeting amyloid, tau, metabolic, and synaptic plasticity.

2.3.4.2 Expert Panel Subgroup Response

While the subgroup found the synthetic control use case interesting, there were several challenges including implementation for certain studies (e.g., pharmaceutical), technical (i.e., need for statistically generated controls) and ethical concerns requiring guidelines for proper use. The expert panel subgroup cautioned against the development of this use case without significant additional scientific development and validation.

2.3.4.3 Full Expert Panel Q&A

<u>Feasibility of Synthetic Controls</u>

The panelists discussed the feasibility of researchers using synthetic control as a comparison group to real-world patients in clinical trials. The expert panel agreed that synthetic controls (also known as counterfactual analysis or digital twins) may be better suited for hypothesis generation as part of target trial design and explainable AI, rather than evidence generation, particularly regarding patient populations and associated risk profiles. Although synthetic controls can aid hypothesis development and eliminate the ethical dilemma of using untreated patient groups in clinical trials, these controls may introduce many confounding factors when compared to real-world patients.

Implementing data on mechanisms of action (MOAs) can help researchers create synthetic controls that better reflect real-world patients, but many AD therapeutics currently in clinical trials lack defined MOAs. In addition, misdiagnosis and mixed dementia captured by ICD codes can further introduce inaccuracy in synthetic controls. The panelists suggested exercising caution when developing CDEs for creating synthetic controls and considered potentially removing synthetic controls as a use case.

<u>Capturing Relevant Data on Caregivers</u>

Caregiver health outcomes can be relevant to studying the outcomes of patients. For example, many interventions are administered to patients by caregivers. If caregivers have health issues, then the administration of interventions to patients may be suboptimal. In addition, patient-reported outcomes (PROs) are often captured through clinicians interacting with caregivers. Some PROs include mental health assessments, such as the Patient Health Questionnaire-9 (PHQ-9), that are intended to be self-administered and not capture the opinion of an outside party. However, EHRs cannot capture information about individuals other than the patient and payer claims data cannot indicate whether the patient or a proxy paid. Panelists suggested that researchers analyze claims data from patients who are partnered or married because the majority of caregivers are spouses.

2.3.4.4 Use Case 3 Future Directions

- Develop new CDEs to support drug approvals that allow use in synthetic control-based trials
- Implement methods to link claims data between patients and their spouses as a first step toward greater linkage of patient and caregiver in RWD

- Consider enhancing payment for telehealth especially telephone-based care to gather addition data from caregivers in claims data

2.4 Session 4: Developing Recommendations

2.4.1 Additional CDEs for AD/ADRD Research Feedback

2.4.1.1 Additional Data Types and Limitations

The final set of presentations related to data elements and types of data that were not covered in the use cases that may require CDE development. Additional data examples, including medical history data, cognitive assessment, and social history, all present opportunities for new data development. Additional data also could include data not collected or recorded but that would be useful for research and clinical purposes but not currently made available to researchers, domains for the raw data from assays versus results (outcomes), and data from domains outside of health and healthcare data. Missing data elements and other limitations could include incomplete, missing, or inaccurate claims information, unstructured, incomputable EHR data, and clinical data of patients pre and post enrollment.

2.4.1.2 Expert Panel Subgroup Response

Panelists discussed challenges in capturing data on MCI, which is relevant to research on tracking the progression of MCI to AD in individuals. Historically, MCI-specific ICD codes are underutilized by clinicians, who often hesitate to officially diagnose patients with MCI because of the heterogeneity of symptoms and phenotypes associated with cognitive impairment. In addition, many medications are associated with the onset of cognitive impairment, which introduces more hesitancy in diagnosis and the use of MCI codes. Because of this hesitancy to diagnose MCI, researchers may use data from individuals likely to have undiagnosed MCI in control groups, which further confounds RWD studies using ICD codes.

2.4.1.3 Full Expert Panel Q&A

<u>Missing Data and Under-reporting Issues</u>

Missing data was discussed. As an alternative to MCI diagnosis codes, other RWD sources can help researchers identify possible MCI in individuals and develop probabilistic flags for undiagnosed patients. First, ICD codes for symptoms of dementia (e.g., amnestic aphasia) can be used as indicators of undiagnosed MCI. In addition, EHRs often contain results from cognitive and functional assessments, which can be linked with dementia symptom ICD codes to potentially identify undiagnosed MCI. However, in the absence of data from imaging or cognitive assessments such as the MoCA (Montreal Cognitive Assessment) or MMSE (Mini-Mental State Examination) by primary care providers and others, diagnosis may not be possible regardless of how data is shaped.

Imaging data may also provide evidence of MCI in patients that did not undergo cognitive and functional assessments. However, researchers should be aware of selection bias when using this approach. Clinicians typically order imaging procedures or assessments for patients who are already suspected of having cognitive impairment, thus reducing the number of individuals undergoing imaging with no suspected cognitive impairment.

The underreporting of MCI in the EHR leads to selection bias. Selection bias leads to clinicians failing to order imaging or neuropsychiatric test unless they believe there is a problem going on. Systematically linking data sources (e.g., cohort and claims) with probabilistic flags to can be used to identify patients with missing MCI diagnoses. Linking cohort and claims data is particularly useful given the regular interval of cohort data available.

<u>Social Determinants of Health and Economic Data</u>

Social determinants of health (SDoH) data are often limited to zip code and must be better captured. Area deprivation index (ADI) data may be more granular and focused. Additional needed information includes structured and unstructured data on the annual wellness visit from the EHR which is not available in claims but is critical to early disease state pharmaceutical interventions. The panel also discussed the need for functional assessment data and better measures of education.

The discussion centered on similar issues of creating CDEs once for use by all researchers and adding flag data to make data more usable discussed as part of the use case 1 discussion. Economic and social information about the patient such as low income, zip code, and dual eligibility may present an opportunity for linkage. For example, the linkage of spouses in claims data using zip codes. Creation of universal CDEs with added flag data to increase the utility was discussed and as part of use case 1.

2.4.1.4 Data Needs Future Directions

- Prioritize diagnosis, prediction, and research into the prevalence and impact of MCI
- Link data and provide cohorts to follow patients for dementia and MCI diagnosis over time
- Enhance the use of SDoH data including but not limited to location (ZIP code and beyond) in RWD
- Research the validity of MCI diagnosis in claims, EHR, and combined data

2.4.2 Overall Additional Feedback

The final session concluded with a description of summary areas for feedback from the panel. These areas were:

1. Propose specific domains for AD/ADRD data
2. Suggest additional CDEs for RWD that may be useful as it pertains to AD/ADRD research
3. Propose new research questions that can be answered using harmonized RWD for NIA's consideration

<u>Domains for Social Context Factors and Biomarkers Data</u>

Panelists also emphasized developing CDEs under all proposed domains from data on nursing home settings and caregivers, such as caregiver demographics and health outcomes. Linking data on caregivers to patient data may help researchers understand how care planning, care environments, and caregiver health can affect patient outcomes. In addition, CDEs can be developed from minimum data sets provided by the Centers for Medicare & Medicaid Services

(CMS), which implements standardized assessments and facilitates care management in nursing homes.

Panelists also suggested expanding the genomics CDE domain to include all relevant AD/ADRD biomarkers, such as data from other "-omic" analyses (e.g., metabolomics), as well as pathology and imaging. However, panelists emphasized prioritizing CDE development in this domain by the frequency that biomarker tests are administered in the clinical setting. For example, data from genetic tests are more commonly available than blood and CSF biomarker analyses; thus, developing CDEs from genetic test results should remain a top priority within this domain.

Missing Data Elements from RWD Sources

In addition to modifying domains, panelists suggested addressing missing data elements from RWD sources to help researchers capture CDEs relevant to top AD/ADRD research questions. Linking data from different RWD sources, such as using EHR data to address missing data elements from claims, can help researchers capture relevant data elements. In addition, researchers can leverage clinical data from research studies focused on underrepresented populations, which have historically been excluded in AD research. Panelists also agreed on prioritizing the development of CDEs that help researchers understand and capture MCI diagnoses through cognitive and functional assessments and imaging data.

Comorbidities Associated with AD/ADRD

Panelists considered how to better define and identify comorbidities that may be associated with AD/ADRD diagnoses. Panelists suggested mapping data on comorbidities to three categories: mental, cognitive, and physical health. These categories can enable researchers to distinguish between comorbidities that impact cognitive decline compared to comorbidities that impact a patient's ability to physically visit a hospital or clinic. In addition, panelists suggested defining how comorbidities are analyzed in longitudinal studies (e.g., identifying timepoints in which comorbidities are relevant in analysis).

Panelists provided overall feedback and priorities for CDE development for AD/ADRD research based on discussions during the meeting, including social context and biomarker domains, missing data, underrepresented populations, MCI detection and data, and comorbidities.

2.4.2.1 Full Expert Panel Round Robin

After a review of the CDE domains and alignment with three use cases, the expert panel provided feedback in a round robin format. This was then summarized into four different themes: data domains, CDEs, new research questions, and additional types of data (see Table 2-1, Table 2-2, Table 2-3, and Table 2-4 below):

Table 2-1. Round Robin Expert Feedback – Data Domains

Comments	Recommendation
• Settings of care • Caregivers and caregiver data • Social risk factors, SDoH and location (ZIP code or urban vs. rural) data	• Add contextual factors or non-health metrics data domain • Explicitly incorporate caregivers
• Genetic testing • Genetics and biomarker data, CSF testing	• Combine these into one domain

Table 2-2. Round Robin Expert Feedback – CDEs

Comments	Recommendation
• Determine update frequency	• Annual or biannual updates • Address the irregular time interval of available data
• New CDEs	• Quality of life (QoL) • End of life care planning • Education, behavioral symptoms • Physician characteristics • Prioritize given the amount of work involved

Table 2-3. Round Robin Expert Feedback – New Research Questions

Comments	Recommendation
• Early diagnosis and treatment of MCI • Diagnosis using harmonized RWD • Risk factor studies and genetic risk factors may be useful • AI as a complement to CDEs	• Use Health and Retirement Study (HRS) observations to come up with validity using AI algorithms • Examine dementia and MCI disease trajectory • Identify and explain regional race and ethnic differences- with regards to subtypes of dementia and diagnostic accuracy and lack thereof

Table 2-4. Round Robin Expert Feedback – Additional Types of Data

Comments	Recommendations
• Functional assessment	• Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) • Separate dementia from MCI and encourage severity assessment.
• Underrepresented populations	• Payments for care other than claims (FQHC, safety net hospitals)
• Linked data	• Medicare and Medicaid (dual eligible) considerations • Combine fee-for-service and Medicare Advantage • Harmonization at a broad level with code books and support for individual researchers

2.4.2.2 Full Expert Panel Additional Discussion

The expert panel also discussed additional issues not covered in the round robin discussion. Those included the Nursing Home Minimum Data Set (MDS), containing longitudinal data not available in claims and EHR data. MDS includes facility characteristics and represents severe disease but does not include length of admission or readmission. There is a need to link nursing home and Medicare claims to address new research.

No best practices were identified to capture minority and underrepresented populations, but consensus was reached on the need for SOGI data and U.S. Census categories. Additional specificity on co-morbidities, particularly mental health, may be important to differentiate between mental health comorbid diagnoses related to AD/ADRD vs. a lifetime condition.

Mapping and harmonization of diagnosis codes over time would be valuable. Outcomes such as caregiver fatigue is important (see Table 1-2). Overall, synthesizing codes over time is a key form of data quality control that includes data semantics and diagnosis over time (longitudinally).

Other areas of feedback provided by panelists included:

- Updating datasets in accordance with the current usage of codes and specifying how codes have been previously implemented in diagnostics.
- Evaluating the benefits and risks of using synthetic controls in clinical trials in which traditional placebo groups may not be feasible.
- Implementing AI algorithms (e.g., ML and NLP) in AD/ADRD research to develop estimates of dementia status by analyzing CDEs from cognitive assessments, functional assessments, and comorbidities.
- Combining fee-for-service and Medicare claims data and tracking plan availability and benefits over time.
- Considering how different data models and initiatives harmonize and map data from different RWD sources and formats

2.4.2.3 Final Panelist Future Directions

The final content-based section concluded with a summary of panelist recommendations, some of which were discussed during the round robin (see Table 2-1, Table 2-2, Table 2-3, and Table 2-4). Those were:

- Domains
    - Adding a CDE domain on social context factors and adding biomarkers to the genomics CDE domain.
    - Capturing relevant data (e.g., health outcomes, demographic data) from caregivers within proposed CDE domains.
    - Seven domains: 1) patient and caregiver information, 2) disease characterization, 3) health assessment, 4) biomarkers and genomics, 5) treatment, 6) patient and caregiver outcomes, and 7) non-health data
    - Include non-health data and contextual factors as well as activities of daily living (ADL) and behavioral / psychiatric symptoms
    - Comorbid diagnoses should include mental health and cognitive specific mappings
- New CDEs
    - Advance care planning
    - Clinician characteristics, practice site
    - Symptom progression

- ZIP code or census tract
- Diagnosis of MCI and prioritizing MCI detection
- Cognitive and functional assessments
- Imaging data

- Research questions
  - What are the best methods for early diagnosis of dementia?
  - How can MCI be detected and how can diagnosis be validated?
  - Assessment of caregiver and non-health (e.g., financial) impact of AD/ADRD
  - Validation of data quality and development of predictive algorithms within and across data sets
- Additional areas of feedback
  - Determine how to use NLP and AI for assessment
  - Clean and synthesize codes that change over time
  - Implement and improve data linkage; Address missing data elements by linking data from EHR text to payer claims data.
  - Leverage data from studies focused on underrepresented populations.
  - Develop consistent semantics
  - Enhance specificity of diagnoses
  - Consider how different data models and initiatives harmonize and map data from different RWD sources and formats.

2.4.3 Closing Remarks

The final session included a review of the discussion during the day.

The participants were thanked for their attention and feedback. A reminder that the meeting outcomes will be posted on the NIA website and that NIA also may choose to publish other materials such as a public facing blog post about the meeting was provided. This ended opportunities to provide input on the meeting topics.

At this point, NIA staff adjourned the meeting.

3 References

1. National Center for Advancing Translational Sciences (NCATS) Toolkit for Patient-Focused Therapy Development - Common data element (CDE). https://toolkit.ncats.nih.gov/glossary/common-data-element-cde/. Accessed 6 Feb 2023

2. U.S. Food and Drug Administration (FDA) Real-World Evidence. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence. Accessed 27 Dec 2022

This page intentionally left blank.

Appendix A Meeting Participants

A.1 Presenters

A.1.1 NIA Division of Behavioral and Social Research

Partha Bhattacharyya, *Chief Data Officer, Office of Data Resources and Analytics (ODRA) and Program Director*

A.1.2 The MITRE Corporation

Allen Leavens, *Principal, Health & Life Sciences*

Robert Lieberthal, *Principal, Health Economics*

Alex Whittaker, *Senior, Health Analytics*

A.2 Expert Panel Members

Matthew Alcusky, *Assistant Professor in the Division of Epidemiology, Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School*

Julie Bynum, *Margaret Terpenning Collegiate Professor of Internal Medicine Professor, Geriatric and Palliative Medicine, Associate Director, Health Policy & Research, Geriatric Center Research Scientist, Institute of Gerontology, University of Michigan*

Jerry Gurwitz, *Dr. John Meyers Professor in Primary Care Medicine, University of Massachusetts Chan Medical School*

Timothy Hohman, *Associate Professor of Neurology, Vanderbilt University Medical Center*

Bryan James, *Associate Professor, Department of Internal Medicine, Rush University Medical College*

Ying-Yee Kong, *Aging Scientific Program Manager, Office of Research & Development, Department of Veterans Affairs*

Jennifer Moser, *Associate Director, Scientific Programs, Million Veteran Program, Office of Research & Development, Department of Veterans Affairs*

Lauren Nicholas, *Associate Professor Department of Health Systems, Management & Policy, CU Anschutz Colorado School of Public Health*

Ana Quiñones, *Associate Professor, OHSU-PSU School of Public Health, Associate Professor of Family Medicine, School of Medicine, OHSU*

David Reuben, *Director, Multicampus Program in Geriatrics Medicine and Gerontology and Chief, Division of Geriatrics, University of California, Los Angeles (UCLA) Center for Health Sciences*

Debby Tsuang, *Director, Puget Sound GRECC, Professor, Department of Psychiatry and Behavioral Sciences, University of Washington*

Kristine Yaffe, *Scola Endowed Chair and Vice Chair, Professor of Psychiatry, Neurology, and Epidemiology, and Director of the Center for Population Brain Health, University of California, San Francisco*

Deriving Common Data Elements from Real-World Data for Alzheimer's Disease and Alzheimer's Disease Related Dementias

Qing Zeng, *Director, Biomedical Informatics Center, Professor, Department of Clinical Research and Leadership, George Washington University School of Medicine and Health Sciences*

Julie Zissimopoulos, *Professor, Sol Price School of Public Policy, Co-Director, Aging and Cognition Program, Schaeffer Center, Co-Director, CeASES ADRD and AD-RCMAR, University of Southern California*

Appendix B Additional Potential CDEs

Table B-1. Potential CDEs for Use Case 1 (Combining Private and Public Payer Data)

Data Types	Examples
AD/ADRD outcomes	**ICD-10**: Alzheimer's (G30.0-G30.9) (F00), Vascular Dementia (F01), Unspecific Dementia (F03.90-F03.91), Mild cognitive impairment (G31.84), Other specified degenerative nervous system disease (Lewy body) (G31.8) **HCC**: 51-52
AD/ADRD co-morbidities	**ICD-10**: Osteoarthritis (M15-M19), Cardiovascular Disease (I51.9), Diabetes (E11.9), Depression (F32.A) **NDC**: Insulin (49502-393)
Cognitive Impairment Screening	**CPT/HCPCS**: Cognitive assessment and care plan services (99483)
Genetic testing for AD	**CPT/HCPCS**: 81401, 81405, 81406
Diagnostics	**CPT/HCPCS**: Positron Emission Tomography scans (78608), spinal puncture (62270)

Table B-2. Potential CDEs for Use Case 2 (Developing CDEs from RWD for Non-Pharmaceutical Intervention Studies)

Data Types	Examples
Common AD/ADRD medications (used to determine current treatment)	• Donepezil (Aricept), e.g., NDCs 0615-7951, 0615-8313 • Memantine (Namenda), e.g., NDCs 0456-3205
Outcomes of focus	• Clinical outcomes-improved cognitive function, diagnosis of disease (MCI, dementia) • Patient Reported Outcomes – ADLs, QoL assessments

Table B-3. Potential CDEs for Use Case 3 (Synthetic Control Methods for AD/ADRD Research using CDE-based RWD)

Data Types	Examples
Basic patient characteristics	Demographics, age
Disease state – Diagnosis	Dementia type, MCI
Clinical outcomes	Cognitive scores over time, amyloid biomarkers
Genomics	Status of Alzheimer's associated genes (APOE status, PSEN1, PSEN2, APP)
Comorbidities (especially those that might prevent study inclusion)	HIV diagnosis, history of transient ischemic attacks (TIA), stroke, or seizures, pregnancy

Appendix C Agenda

C.1 Title

Deriving Common Data Elements from Real-Word Data for Alzheimer's Disease and Alzheimer's Disease Related Dementias (AD/ADRD) — Technical Expert Panel Meeting

C.2 Date

February 6, 2023, 10:00 a.m. to 4:00 p.m., Eastern Time

C.3 Location

Zoom meeting: https://tinyurl.com/bdfyu34k

C.4 Purpose and Background

The National Institute on Aging (NIA) is convening an exploratory discussion to identify opportunities to accelerate Alzheimer's Disease and Alzheimer's Disease Related Dementias (AD/ADRD) research using real-world-data (RWD) sources. The meeting will focus primarily on Common Data Element (CDE) methods that can be applied for harmonizing data contained in RWD including healthcare claims and electronic health records (EHRs). The meeting output will include the expert panel's recommendations to NIA on the applicability of CDEs to AD/ADRD research in the form of formal meeting notes.

C.5 Meeting Schedule

10:00 – 10:30 am: Session 1 | Introduction and Purpose

- **Opening Remarks:** NIA Leadership
- **Meeting Purpose:** Partha Bhattacharyya
- **Pre-meeting Poll Results:** MITRE

The opening session is designed to set the stage for meeting purpose and introduce methods for identifying and defining common data elements (CDE) when using RWD. This session also will include results of a poll regarding gaps, opportunities for research and questions in advance of the meeting.

Objectives

- Establish the call to action
- Review pre-meeting feedback from panelists

10:30 – 11:45: Session 2 | CDE Development

- **CDE Models:** MITRE
- **Q&A:** Expert Panel
- **CDE Domains:** MITRE
- **Initial Domain Feedback:** CDE subgroup of expert panel

The CDE development portion of the meeting will focus on the overall approach to harmonizing RWD through existing CDE models (US Core Data for Interoperability (USCDI), National Institute of Neurological Disorders and Stroke Common Data Elements (NINDS CDEs), HL7 Fast Healthcare Interoperability Resources (FHIR), and the Observational Medical Outcomes Partnership (OMOP) Common Data Model). The entire expert panel then will have the opportunity to ask questions regarding the models described in the pre-reading material and presented at the meeting. A presentation of proposed domains (patient information, disease characterization, health assessment, genomics, treatment, outcomes) for AD/ADRD CDEs then will be described. A subgroup of the entire panel with specific expertise in CDEs then will be asked to comment individually on the CDE domains, needed refinements, and preliminary recommended changes to the presented CDE domains.

Objectives

- Assess expert perspective on existing CDE models
- Present domains for AD/ADRD CDEs
- Elicit preliminary expert feedback on the CDE domains for data harmonization

11:45 –12:30: Lunch Break

12:30 – 2:15: Session 3 | Use Cases

- **Use Case 1 – Combining private insurer and public payer data:** MITRE
  - **Discussion:** Use Case 1 subgroup of expert panel
- **Use Case 2 – Developing CDEs from RWD for non-pharmaceutical intervention studies:** MITRE
  - **Discussion:** Use Case 2 subgroup of expert panel
- **Use Case 3 – Synthetic control methods for AD/ADRD research using CDE-based RWD:** MITRE
  - **Discussion:** Use Case 3 subgroup of expert panel

The focus of the discussion will move to specific areas where the development of specific CDEs derived from claims and EHR data are likely to have positive impact. Three use cases will be presented for consideration by the expert panel. For each use case, an identified subgroup of the overall expert panel will have the chance to provide their perspective on the prospects for defining CDEs in each use case, as well as preliminary recommendations and next steps for implementation.

Objectives

- Assess research questions that can be achieved using RWD with CDEs
- Determine a priority list of next steps for CDE development that can be achieved
- Generate actionable research priorities for implementing each use case

2:15 –2:30: Short Break

2:30 – 3:45: Session 4 | Developing Recommendations

- **Summary of Areas for Feedback:** MITRE
- **Overall CDE Domain Feedback:** Expert panel discussion
- **Recommendations Discussion:** Expert panel round robin

The final section of the meeting will elicit overall feedback from participants on the identified set of CDE domains along with a broader set of recommendations including deriving CDEs from RWD and NIA research opportunities. Participants will have the opportunity to 1) propose specific adjustments to the preliminary domains for AD/ADRD data including new domains or consolidating domains, 2) suggest additional CDEs for RWD, and 3) propose new research questions that can be answered using harmonized RWD for NIA's consideration.

Objectives

- Revise the proposed CDE domains based on expert feedback
- Determine additional CDEs for AD/ADRD research that should be developed
- Identify research questions that would advance the field of aging research

3:45 – 4:00: Session 5 | Closing

- **Closing and Next Steps:** MITRE

MITRE staff will summarize meeting highlights. The meeting findings will be posted on NIA's website.

Appendix D Abbreviations and Acronyms

Term	Definition
ADI	Area deprivation index
ADL	Activities of daily living
AD/ADRD	Alzheimer's Disease and Alzheimer's Disease Related Dementias
AI	Artificial intelligence
CDE	Common data element
CDM	Common data model
CMS	Centers for Medicare and Medicaid Services
CPT/HCPCS	Current Procedural Terminology (CPT) / Healthcare Common Procedure Coding System (HCPCS) codes
CSF	Cerebrospinal fluid
EHR	Electronic health record
FFS	Fee-for-service
FHIR	Fast Healthcare Interoperability Resources
HCC	Hierarchical Condition Category code
IADL	Instrumental activities of daily living
ICD-10	International Classification of Diseases, 10th Revision
MA	Medicare Advantage
MCI	Mild cognitive impairment
MDS	Minimum data set
MMSE	Mini Mental State Examination
MoCA	Montreal Cognitive Assessment
NACC	National Alzheimer's Coordinating Center
NDC	National Drug Code
NIA	National Institute on Aging
NIH	National Institutes of Health
NINDS	National Institute of Neurological Disorders and Stroke
NLP	Natural language processing
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PET	Positron emission tomography
PHI	Protected health information

PII	Personally identifiable information
PRO	Patient reported outcome
QoL	Quality of life
RELD / SOGI Data	Race, ethnicity, language, and disability status (RELD) and sexual orientation and gender identity (SOGI) data
RWD	Real-world data
SDoH	Social determinants of health
USCDI	United States Core Data for Interoperability
VA	Department of Veterans Affairs